

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 14:59:26 ; Search time 34.67 seconds  
(without alignments)  
1681.965 Million cell updates/sec

Title: DAVISN-P04196.PPP  
Perfect score: 3036  
Sequence: 1 MKALIAALLITLOYSCAVS.....GKFKSGPQVSMFTTFPPK 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A_Geneseq_032802.*			
1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323.5	10.7	245	22	ABG28019 Novel human diagno
2	318.5	10.5	330	22	ABG25331 Novel human diagno
3	311	10.2	491	22	ABG09949 Novel human diagno
4	306	10.1	351	13	RAA24393 Sequence of Histid
5	296.5	9.8	406	22	ABG27250 Novel human diagno
6	284.5	9.4	233	22	ABG19767 Novel human diagno
7	283.5	9.3	292	22	ABG32108 Peptide #4759 enco
8	283.5	9.3	292	22	ABG37360 Peptide #4866 enco
9	283.5	9.3	292	22	ABG22646 Protein #4645 enco
10	283.5	9.3	292	22	AAW58050 Human brain expres
11	283.5	9.3	292	22	AAW70490 Human bone marrow

12	283.5	9.3	292	22	AAW18326 Peptide #4760 enco
13	283.5	9.3	292	22	AAW30814 Peptide #4851 enco
14	283.5	9.3	292	22	AAW05933 Peptide #4615 enco
15	283	9.3	183	22	ABG25337 Novel human diagno
16	273.5	9.0	598	22	ABG23595 Novel human diagno
17	263	8.7	324	22	ABG20509 Novel human diagno
18	256	8.4	594	22	ABG1362 Drosophila melanog
19	251.5	8.3	227	22	ABG14399 Novel human diagno
20	250.5	8.3	304	22	ABG08412 Novel human diagno
21	234	7.7	124	22	ABG12879 Novel human diagno
22	226.5	7.5	388	22	ABW51345 Mouse HS-glycoprot
23	224.5	7.4	1401	22	ABG24037 Novel human diagno
24	224	7.4	382	20	AAW88491 Human liver clone
25	224	7.4	382	21	AAW25782 Human secreted pro
26	224	7.4	382	22	AAW51346 Human HS-glycoprot
27	224	7.4	382	22	AAW75368 Human secreted pro
28	222.5	7.3	372	22	ABG67014 Drosophila melanog
29	217.5	7.2	167	22	ABG64598 Drosophila melanog
30	217	7.1	266	22	ABG03494 Novel human diagno
31	217	7.1	485	22	ABW59608 Drosophila melanog
32	217	7.1	644	22	ABG21101 Novel human diagno
33	213.5	7.0	102	22	ABG68218 Drosophila melanog
34	211	6.9	449	22	ABW71938 Drosophila melanog
35	210	6.9	82	22	ABW27363 Human peptide #14
36	210	6.9	82	22	ABW32509 Peptide #15 encode
37	210	6.9	82	22	ABW18015 Protein #14 encode
38	210	6.9	82	22	AAW53343 Human brain expres
39	210	6.9	82	22	AAW65721 Human bone marrow
40	210	6.9	82	22	AAW13580 Peptide #14 encode
41	210	6.9	82	22	AAW25978 Peptide #15 encode
42	210	6.9	82	22	AAW01333 Peptide #15 encode
43	207.5	6.8	79	22	AAU32905 Novel human secret
44	204.5	6.7	378	22	AAW51347 Bovine HS-glycopro
45	202.5	6.7	940	22	AAU28194 Novel human secret

ALIGNMENTS

RESULT 1	
ABG28019	ABG28019 standard; Protein; 245 AA.
ID	ABG28019;
XX	AC
XX	ABG28019;
DT	18-FEB-2002 (first entry)
XX	Novel human diagnostic protein #28010.
DE	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder.
KW	Homo sapiens.
XX	WO200175067-A2.
PN	11-OCT-2001.
XX	30-MAR-2001; 2001WO-US08631.
PF	31-MAR-2000; 2000US-0540217.
XX	23-AUG-2000; 2000US-0649167.
PR	(HYSE-) HYSEQ INC.
XX	Drmanac RT, Liu C, Tang YT;
PA	WPI; 2001-639362/73.
XX	N-PSDB; AAS92206.
DR	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	

```
PT biodiversity
XX Claim 20; SEQ ID No 58378; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 245 AA;
XX
XX Query Match 10.7%; Score 323.5; DB 22; Length 245;
XX Best Local Similarity 30.0%; Pred. No. 7.2e-21;
XX Matches 76; Conservative 15; Mismatches 107; Indels 55; Gaps 9;
XX
XX QY 258 HLGHFHWGHERSSTTKPPFKPHGSRDHHHPKHEHSGPPPPDERDHSGLPQGGP 317
XX Db 21 hyhhhyhhhhhh-----hyhrhhhhphhhphhhhhhhhhhhhhhhhhhh 71
XX
XX QY 318 PLLPMSCSCQATFGTNGAQRSHNNSSDLPHPKHHSHEQHPHGHAPH-----HPHE 373
XX Db 72 -----hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 109
XX
XX QY 374 HDTHRQHPHGHHPHGHHPHGHHPH-----GHHP-----HCHDFQDYGCDPPPHNQHCCH 429
XX Db 110 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh--h 167
XX
XX QY 430 GHGPPGHLRRRGPGKPRPFHCRQIGSVYRLPLPKRGEVLPLPEANFPSPFLPH-----484
XX Db 168 hhhphhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh--phraayl 224
XX
XX QY 485 -----HKHPLKP 491
XX Db 225 lglsnfrkpfpr 237
XX
XX RESULT 2
XX ABG25331
XX ID ABG25331 standard; Protein; 330 AA.
XX
XX AC ABG25331;
XX
XX DE 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #25322.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
```

```
PF PF 30-MAR-2001; 2001WO-US08631.
XX XX
PR PR 31-MAR-2000; 2000US-0540217.
PR PR 23-AUG-2000; 2000US-0649167.
XX XX
PA (HYSE-) HYSEQ INC.
XX XX
PI Drmanac RT, Liu C, Tang YT;
XX XX
DR WPI; 2001-639362/73.
DR DR N-PSDB; AAS89518.
XX XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 20; SEQ ID No 55690; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 330 AA;
XX
XX Query Match 10.5%; Score 318.5; DB 22; Length 330;
XX Best Local Similarity 29.6%; Pred. No. 3e-20;
XX Matches 80; Conservative 12; Mismatches 99; Indels 79; Gaps 11;
XX
XX QY 249 HENINGVPPHLGHPF-HWGHERSSTTKPPFKPHGSRDHHHPKHEHSGPPPPDERDHS 307
XX Db 124 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh--hph- 180
XX
XX QY 308 HGPPLPQGGPPPLLPMSCSQATFGTNGAQRSHNNSSDLPHPKHHSHEQHP-HGHHP 366
XX Db 181 hqp-----qhhhhq-----yhhhyhhhhhhhhhhhhhhhh 209
XX
XX QY 367 HAHHPHEHDTROHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 426
XX Db 210 hhyhyhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh--phnhyh 259
XX
XX QY 427 CCHGHP-----PPGHLRRRG-----GKPRPFHCRQIGSVYRLPLPKRGEVLPLPEANF 477
XX Db 260 --hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh-- 300
XX
XX QY 478 PSEPLPHKHLKPDNQPFQSVSESCPGK 507
XX Db 301 -----hhhyhtssvgpgsqpagaslpqg 325
XX
XX RESULT 3
XX ABG09949
XX ID ABG09949 standard; Protein; 491 AA.
XX
```



Query Match 10.1%; Score 306; DB 13; Length 351;  
Best Local Similarity 34.0%; Pred. No. 4.4e-19;  
Matches 66; Conservative 10; Mismatches 86; Indels 32; Gaps 8;

QY 246 POEHENINGVPHLGHPPFWGCHERSSTTKPFKPHGSRDHHH-PHKPHGHPPPPPDER 304  
DB 98 phphhhhhhhphh--hhhhghhhhh-----haahhhheehhhhhhaahhhheeh 147  
QY 305 DSHGPPPLPQGPPLLPMSCSQCATFTGNGAQRSHNNNSDLHPKHHSHEQHPHGH 364  
DB 148 hhhhhhaahhp-----wfhhlghyhhhaaphhhhhhhha---phhhhhhhaphhh 197  
QY 365 H-----PHAHHPHEH--DTHRHQPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 417  
DB 198 hnhhhaphhhhhhhhphhhhhhhhhghhhhhhhhhghhhhhhhhhghhhhhhhhhhd---- 253  
QY 418 DEPPHNOGHCCHGH 431  
DB 254 ahhhhhhhhdahhh 267

RESULT 5  
ID ABG27250 standard; Protein; 406 AA.  
XX AC ABG27250;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #27241.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS91437.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 57609; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 406 AA;

Query Match 9.8%; Score 296.5; DB 22; Length 406;  
Best Local Similarity 33.7%; Pred. No. 3.8e-18;  
Matches 59; Conservative 5; Mismatches 64; Indels 47; Gaps 4;

QY 276 PFKPHGSRDHHHPKHHPHGHGPPPPPPDERDHSHPPLPQGP----- 317  
DB 139 pppsp-----spppppppppppppppppppppppppppppppppppsp 191  
QY 318 -PLPMSCSQCATFTGNGAQRSHNNNSDLHPKHHSHEQHPHGHHPHHPHEHDT 376  
DB 192 spslppsp-----spphhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 242  
QY 377 HRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 431  
DB 243 hhh 285

RESULT 6  
ID ABG19767 standard; Protein; 233 AA.  
XX AC ABG19767;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #19758.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS83954.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 50126; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:01:21 ; Search time 15.34 Seconds  
(without alignments)  
835,948 Million cell updates/sec

Title: DAVISM-P04196.PEP  
Perfect score: 3036  
Sequence: 1 MKALIALLLITLQYSCAVS.....GPKSGPPQVSMFFTHFPK 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217.5	7.2	167	4	US-09-507-323B-13
2	198.5	6.5	796	4	US-08-857-076-40
3	187.5	6.2	858	4	US-08-857-076-41
4	187.5	6.2	892	4	US-08-857-076-42
5	185	6.1	26	3	US-08-785-636-1
6	185	6.1	26	3	US-09-095-407-2
7	169	5.6	1185	4	US-09-041-886-23
8	165.5	5.5	309	1	US-08-161-406-2
9	160.5	5.3	1958	1	US-07-945-283-2
10	157.5	5.2	189	1	US-08-152-922A-6
11	153	5.0	89	4	US-09-507-323B-12
12	147.5	4.9	306	1	US-08-217-327-6
13	144.5	4.8	60	1	US-08-255-457-1
14	144.5	4.8	60	2	US-09-115-032-1
15	144.5	4.8	60	5	PCT-US95-0572-1
16	141.5	4.7	285	1	US-08-292-045-7
17	137.5	4.5	285	1	US-08-292-045-5
18	134.5	4.4	313	3	US-08-686-528A-3
19	134.5	4.4	313	4	US-09-456-287-3
20	134.5	4.4	337	3	US-08-686-528A-2
21	134.5	4.4	337	4	US-09-456-287-2
22	134	4.4	398	4	US-09-461-474-17
23	129.5	4.3	402	4	US-09-029-755C-5
24	128.5	4.2	474	4	US-09-461-474-10
25	126.5	4.2	285	1	US-08-292-045-10
26	126	4.2	552	4	US-09-219-849-7
27	126	4.2	960	4	US-09-219-849-6

28	125.5	4.1	255	4	US-09-612-126-1	Sequence 1, Appl
29	125.5	4.1	265	1	US-08-292-045-2	Sequence 2, Appl
30	125.5	4.1	509	4	US-08-860-635A-21	Sequence 21, Appl
31	125.5	4.1	509	4	US-09-281-476-21	Sequence 21, Appl
32	125	4.1	349	4	US-09-461-474-12	Sequence 12, Appl
33	125	4.1	960	4	US-09-219-849-5	Sequence 2, Appl
34	124.5	4.1	826	5	PCT-US92-00018-2	Sequence 2, Appl
35	124.5	4.1	1182	4	US-09-041-886-21	Sequence 21, Appl
36	124	4.1	448	4	US-09-461-474-8	Sequence 8, Appl
37	123.5	4.1	507	4	US-08-860-635A-19	Sequence 19, Appl
38	123.5	4.1	507	4	US-09-281-476-19	Sequence 19, Appl
39	123.5	4.1	367	2	US-08-737-045-14	Sequence 14, Appl
40	123	4.1	367	3	US-08-932-871B-2	Sequence 2, Appl
41	123	4.1	367	3	US-09-476-919-2	Sequence 2, Appl
42	123	4.1	367	4	US-08-780-311A-2	Sequence 2, Appl
43	123	4.1	124	2	US-08-925-237-2	Sequence 2, Appl
44	122.5	4.0	124	2	US-08-737-045-12	Sequence 12, Appl
45	122.5	4.0	352	2	US-08-737-045-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-507-323B-13  
; Sequence 13, Application US/09507323B  
; Patent No. 6274345  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Jong Hyun  
; APPLICANT: Lee, Sang Yup  
; APPLICANT: Xu, Zhaohui  
; TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE  
; TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E.COLI AS CELL SURFACE ANCHORING  
; TITLE OF INVENTION: MOTIF  
; FILE REFERENCE: HYLEE39.001AUS  
; CURRENT APPLICATION NUMBER: US/09/507,323B  
; CURRENT FILING DATE: 2000-02-18  
; PRIOR APPLICATION NUMBER: KR 10-1999-0005773  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Histidine linker.  
US-09-507-323B-13

Query Match 7.2%; Score 217.5; DB 4; Length 167;  
Best Local Similarity 30.7%; Pred. No. 3.5e-13;  
Matches 55; Conservative 4; Mismatches 67; Indels 53; Gaps 9;

Qy	280	PHGSRDHHP-----HKPHEHGPPPPPPDERDHSHGPPPLPOGPPPLPMSCSSCQHATFG	333
Db	5	PSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSG-----	46
Qy	334	TNGAQRHSNNNSDLHPKHHSH-----EQHPCHGH-----PHAHPHEHDTROHPHGH	384
Db	47	-----HHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHS--GLDPSGH	99
Qy	385	HPHGH---HPHGHHPHGH---HPHGHHPHGHFDQYGCDDPPHNOGHCCHGHPGPH	437
Db	100	HHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHH--HSGLDPSGH	151

RESULT 2  
US-08-857-076-40  
; Sequence 40, Application US/08857076C  
; Patent No. 6225120  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary

; APPLICANT: Kimura, Koutarou  
; APPLICANT: Patterson, Garth  
; APPLICANT: Ogd, Scott  
; APPLICANT: Paradis, Suzanne  
; APPLICANT: Tissenbaum, Heidi  
; APPLICANT: Morris, Jason  
; APPLICANT: Kowek, Allison  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351001  
; CURRENT APPLICATION NUMBER: US/08/857,076C  
; CURRENT FILING DATE: 1997-05-15  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 40  
; LENGTH: 796  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-08-857-076-40

Query Match 6.5%; Score 198.5; DB 4; Length 796;  
Best Local Similarity 20.4%; Pred. No. 2.1e-10;  
Matches 124; Conservative 74; Mismatches 204; Indels 207; Gaps 30;

QY 4 LIAALLITLOYSKAVSPDTC-----SAVEPEAKALDLINK--RRRDGYLFQLLRIA 54  
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 3 LIATSLVPEDEHTPMSPVNTTKILQSGIKMEIPPYLDPSQDDDPEDG-----VNVP 57  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 55 DAHLDRVENTVYVLDVQESDCSVL--SRKYNDCEPPDSRRPSEIIVIGOCKVIATRH 112  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 58 DPDLFTKNTNTEYDLDVLKLGKPAVDEARK--KIEVPDASAPPNKIVEYLMYRTLK 114  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 113 SHESQDLRVDFNCTTSSVSSALANTKDSPLIDFFEDTERYKQANKALEKYKEENDDF 172  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 115 ESE-----LIQLNAYRTKRNLSLNLVKNKI-----DREEDQKACESLVKLLKDKKNDL 163  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 173 ASFRVDRIEVARVRGEGTGYF-----VDFSVRNCPRHHP----- 209  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 164 QNL-ID-----VLSKGTKYTGCTITPTLDGLQVHGRKGFPHVYVYKGLWRFNEMTKNET 218  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 210 RHPNVGFCRADLYDVEALDLESK-----NLVINCEVDFDQEHENINGVPPHLGH 261  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 219 RHVD---HCK-----HAFEMKSDMVCVNPYHYEIVGTMIVGQDHDNRDMPPH--Q 266  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 262 PFHWGHE-----RSSTTKPPFKPH-----GSRDHHHPHK-PHEH 295  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 267 RYHTPGQDPVDDMSRFIPPAIRPPPMNMHTRPQMPQOLPSVGATFAHPLPHQAPHNP 326  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 296 GPPPPPPDERDHS-----GPPPLQGPPLPPLPMSCCSOHATFTGNGAQRHSHNNN 345  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 327 GVSHPYSIAPQTHYPLNMNPIPOMQPMQMPPL-----HOGYGMNGPSCSENNN 377  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 346 SSDLHPKHHSHEQHPHGHHPHHAHPHEHDTHRQHPHGHHPHGHHPHGHHPHGH 405  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 378 --PFHONHHYNDISHPN-HYSYDCGNLYGFTPTPYDFHHPNQPH----- 421  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 406 PHCHDFQDYGCPDPPHNGCHCHGHPGPPGLHRRRGPKGP--RPF----- 450  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 422 -----QPPLSQNHTSQGSHQPGH--QGQVNDPPIISRPLVLPSTVTLVDVFR 468  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 451 HCRQI-----GSVYR-----LPPLRKGVEVLP--LPEAN 476  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 469 YCQQTGNGRFFEGESQSGAIISSNKFTIEEDSPICGVTVVVRPRMTDGEVLENTIMPE-- 526  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 477 FPSFPLPHH 485  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 527 ----DAPYH 531  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 3  
US-08-857-076-41  
; Sequence 41, Application US/08857076C

; Patent No. 6225120  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Kimura, Koutarou  
; APPLICANT: Patterson, Garth  
; APPLICANT: Ogd, Scott  
; APPLICANT: Paradis, Suzanne  
; APPLICANT: Tissenbaum, Heidi  
; APPLICANT: Morris, Jason  
; APPLICANT: Kowek, Allison  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351001  
; CURRENT APPLICATION NUMBER: US/08/857,076C  
; CURRENT FILING DATE: 1997-05-15  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 41  
; LENGTH: 858  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-08-857-076-41

Query Match 6.2%; Score 187.5; DB 4; Length 858;  
Best Local Similarity 20.2%; Pred. No. 2.7e-09;  
Matches 111; Conservative 63; Mismatches 182; Indels 193; Gaps 27;

QY 55 DAHLDRVENTVYVLDVQESDCSVL--SRKYNDCEPPDSRRPSEIIVIGOCKVIATRH 112  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 120 DPDLFTKNTNTEYDLDVLKLGKPAVDEARK--KIEVPDASAPPNKIVEYLMYRTLK 176  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 113 SHESQDLRVDFNCTTSSVSSALANTKDSPLIDFFEDTERYKQANKALEKYKEENDDF 172  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 177 ESE-----LIQLNAYRTKRNLSLNLVKNKI-----DREEDQKACESLVKLLKDKKNDL 225  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 173 ASFRVDRIEVARVRGEGTGYF-----VDFSVRNCPRHHP----- 209  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 226 QNL-ID-----VLSKGTKYTGCTITPTLDGLQVHGRKGFPHVYVYKGLWRFNEMTKNET 280  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 210 RHPNVGFCRADLYDVEALDLESK-----NLVINCEVDFDQEHENINGVPPHLGH 261  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 281 RHVD---HCK-----HAFEMKSDMVCVNPYHYEIVGTMIVGQDHDNRDMPPH--Q 328  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 262 PFHWGHE-----RSSTTKPPFKPH-----GSRDHHHPHK-PHEH 295  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 329 RYHTPGQDPVDDMSRFIPPAIRPPPMNMHTRPQMPQOLPSVGATFAHPLPHQAPHNP 388  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 296 GPPPPPPDERDHS-----GPPPLQGPPLPPLPMSCCSOHATFTGNGAQRHSHNNN 345  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 389 GVSHPYSIAPQTHYPLNMNPIPOMQPMQMPPL-----HOGYGMNGPSCSENNN 439  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 346 SSDLHPKHHSHEQHPHGHHPHHAHPHEHDTHRQHPHGHHPHGHHPHGHHPHGH 405  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 440 --PFHONHHYNDISHPN-HYSYDCGNLYGFTPTPYDFHHPNQPH----- 483  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 406 PHCHDFQDYGCPDPPHNGCHCHGHPGPPGLHRRRGPKGP--RPF----- 450  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 484 -----QPPLSQNHTSQGSHQPGH--QGQVNDPPIISRPLVLPSTVTLVDVFR 530  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 451 HCRQI-----GSVYR-----LPPLRKGVEVLP--LPEAN 476  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 531 YCQQTGNGRFFEGESQSGAIISSNKFTIEEDSPICGVTVVVRPRMTDGEVLENTIMPE-- 588  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 477 FPSFPLPHH 485  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 589 ----DAPYH 593  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 4  
US-08-857-076-42  
; Sequence 42, Application US/08857076C  
; Patent No. 6225120



```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-095-407-2

Query Match          6.1%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 GHHPHGHHPHGHHPHGHHPH 407
Db 1 GHHPHGHHPHGHHPHGHHPH 25

RESULT 7
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredezen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-23

Query Match          5.6%; Score 169; DB 4; Length 1185;
Best Local Similarity 23.5%; Pred. No. 2.5e-07;
Matches 87; Conservative 27; Mismatches 127; Indels 130; Gaps 21;

QY 255 VPHLGHHPHWG-----GHER-----SSTTKPPFK-PHGRDRHH--- 288
Db 325 LPHGLPSYAMGQGMGLPPGPEKGTLPSPHSLPPASSAPAPMPRPYSSSSSSAA 384
QY 289 -----PHKPHGPPPPPPDERDHSHP-----LP-----OG 315
Db 385 ASSSSSSSSSAPFPASQALPSYHSF---PPPTSLSVSNQPKYQTQPSLPSQAVWSOG 441
QY 316 PPLPLP-----MSCSCQHATF-GTNGAQRHSHNNSSDLPHKHHSHQHPHGHHPH 370
Db 442 PPPPPYGRLLANSNAHPGPPFPSTCAQSTAHPVPST--HHHHHQOQQOQQOQQOQH 499
```

---

```

QY 371 PHEHDTHQHPHGHHPHGHHPH---GHHPHGHHPHGHHPHCHDFQDY--GPCD-PPPHNQ 424
Db 500 GNS-----GPPPPGAFPHPLEGSSHHAHYAMSPSLGSLRYPYPCGPAHLPPHSQ 550
QY 425 GHCCHG--HGPPPGH-----LRRGPKGKPRPFHCRQIGSVYRLPLLRK-- 466
Db 551 VSYSQAGPNPVPVSSSSSSSTSGSYPCSHSPSQGQO-----GAPYPPPPYPTVT 603
QY 467 -----GEVLPPLP---EANFPSPFPLPHHKHPLKPDNQPPQSVSESCPKGKSGFP 513
Db 604 TSSATLSTVIATVASSPAGYKTASPPGPPYGK-----RAPSPGAYKATATPPGYKGP 657
QY 514 QVSMFFTHFTFP 524
Db 658 --PSFRTGTPP 666

RESULT 8
US-08-161-406-2
; Sequence 2, Application US/08161406
; Patent No. 5476785
; GENERAL INFORMATION:
; APPLICANT: Wellens, Thomas
; APPLICANT: Howard, Russell J.
; TITLE OF INVENTION: Recombinant DNA Clone Containng a
; TITLE OF INVENTION: Genomic Fragment of PfHRP-II Gene from Plasmodium
; TITLE OF INVENTION: Falciparum
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/161,406
; FILING DATE: 06-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1173-448P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-161-406-2

Query Match          5.5%; Score 165.5; DB 1; Length 309;
Best Local Similarity 27.3%; Pred. No. 7.9e-08;
Matches 50; Conservative 13; Mismatches 67; Indels 53; Gaps 14;

QY 268 HRSSTTKPPFKPHGRDHHHPKHHEHGGPPPPPPDERDHSHPPLPQGPPLPLPMSCSS 327
Db 25 HETQAVDDAHNAHHVADAHNAHHAA-----DAHHAH----- 59
QY 328 QHATFTNGAQRHSHNNSSDLPHKHHSHQHPHGHHPH-PhaHHPHEHDTHQHPH- 382
Db 60 -HA-----ADAH-HAHAADAH-HAHAADAH-HAHAADAH-HAHAADAHHAAD 108
```









QY 426 HCCHGH 431  
|||||  
Db 53 GCCHGH 58

Search completed: May 31, 2002, 15:03:36  
Job time: 135 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:01:16 ; Search time 20.98 Seconds  
(without alignments)  
2404.523 Million cell updates/sec

Title: DAVISN-P04196.PEP

Perfect score: 3036

Sequence: 1 MKALIAALLILQYSCAVS.....GKFKSGFPQVSMFFTHTPPK 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71.\*

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3036	100.0	525	1 KGHUGH	histidine-rich gly
2	1514.5	49.9	445	2 A60488	histidine-rich gly
3	306	10.1	351	1 KGZQHL	histidine-rich gly
4	269	8.9	735	2 T45059	hypothetical prote
5	266	8.8	140	2 A54523	histidine-rich prote
6	223.5	7.4	471	2 T33997	hypothetical prote
7	217	7.1	644	1 KGHUHI	kininogen, HMW pre
8	213	7.0	774	2 T39539	alpha-amylase homo
9	209.5	6.9	1002	2 T43236	carboxypeptidase C
10	202.5	6.7	1571	2 T14155	zinc finger protei
11	199.5	6.6	834	2 T29821	hypothetical prote
12	197	6.5	330	2 S22140	nodulin Enod2 - Se
13	197	6.5	620	2 S06733	hydroxyproline-ric
14	194.5	6.4	2715	2 T13049	eyelid - fruit fly
15	193.5	6.4	350	2 T55214	salivary proline-r
16	193	6.4	621	1 KGB0H1	kininogen, HMW I p
17	192.5	6.3	1560	2 T00080	hypothetical prote
18	192	6.3	1378	2 T30173	zinc finger protei
19	190	6.3	619	1 KGB0H2	kininogen, HMW II
20	187.5	6.2	940	2 J60291	FB19 protein - hum
21	186	6.1	309	2 S08343	nodulin precursor
22	186	6.1	436	2 T49714	MHC H-2K/t-w5-link
23	186	6.1	508	2 A45477	cAMP response elem
24	184.5	6.1	309	2 S10889	proline-rich prote
25	180	5.9	242	2 H82061	hypothetical prote
26	179.5	5.9	580	2 T10863	extensin precursor
27	179.5	5.9	657	2 A29454	knob-associated hi
28	178.5	5.9	639	2 A25486	kininogen, HMW I p
29	178	5.9	852	2 A34373	histidine-rich cal

ALIGNMENTS

RESULT 1

KGHUGH

histidine-rich glycoprotein precursor - human

N:Alternate names: HRG

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000

C:Accession: A01287; S29669

R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A>Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu

A:Reference number: A01287; MUID:86216149

A:Accession: A01287

A:Molecule type: mRNA

A:Residues: 1-525 <KOI>

A:Cross-references: GB:AB005803; MID:g2280513; PIDN:BAA21613.1; PID:g2280514

R:Hennis, B.; Havelaar, A.; Kluit, C.

submitted to the EMBL Data Library, October 1991

A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly

A:Reference number: S29669

A:Accession: S29669

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 214-247 <HEN>

A:Cross-references: EMBL:Z17218; MID:g32453; PIDN:CAAY8925.1; PID:g32454

C:Comment: Although its physiological function is not yet known, HRG does bind heme, din, and the lysine-binding site of plasminogen. On the basis of its homology with HM

lood coagulation cascade.

C:Comment: The amino half of this protein is homologous to the first two cystatin-lik

could not have inhibitory activity.

C:Comment: In addition to having a high histidine and proline content, this protein h

e-rich' region.

C:Genetics:

A:Gene: GDB:HRG

A:Cross-references: GDB:120055; OMIM:142640

A:Map position: 3q27-3q27

C:Superfamily: histidine-rich glycoprotein; cystatin homology

C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F:19-131/Domain: cystatin homology <CYI>

F:140-246/Domain: cystatin homology <CY2>

F:276-321/Region: proline-rich

F:348-437/Region: histidine-rich

F:351-497/Region: proline-rich

F:63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 100.0%; Score 3036; DB 1; Length 525;

Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKALIAALLLITLQYSCAVPTDCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLD 60  
Db 1 MKALIAALLLITLQYSCAVPTDCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLD 60  
QY 61 VENTTVYLVLDVQSDCVLSRKYNWCEPPDSRRPGEIVIGQCKVIATRHSHESQDLR 120  
Db 61 VENTTVYLVLDVQSDCVLSRKYNWCEPPDSRRPGEIVIGQCKVIATRHSHESQDLR 120  
QY 121 VIDFNCTSSVSALANTKDSPLVLDFFEDTERYRKQANKALEKYEENDDFASFRVDRI 180  
Db 121 VIDFNCTSSVSALANTKDSPLVLDFFEDTERYRKQANKALEKYEENDDFASFRVDRI 180  
QY 181 ERVARVGGEGTYGVDFSVRNCPRHHPNPFVFCRADLFYDVEALDLESPKLNLYN 240  
Db 181 ERVARVGGEGTYGVDFSVRNCPRHHPNPFVFCRADLFYDVEALDLESPKLNLYN 240  
QY 241 CEVFDPOEHENINGVPHLGHFGHGHERSSTTKPKPKGSRDHHHPKPHGHPPPP 300  
Db 241 CEVFDPOEHENINGVPHLGHFGHGHERSSTTKPKPKGSRDHHHPKPHGHPPPP 300  
QY 301 PDERHSHGPPPLPQGPPLPMSCCSCQHATFTNGAQRHSHNNNSDLHPKHSHSQH 360  
Db 301 PDERHSHGPPPLPQGPPLPMSCCSCQHATFTNGAQRHSHNNNSDLHPKHSHSQH 360  
QY 361 PHGHHPHHPHEDTHRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHP 420  
Db 361 PHGHHPHHPHEDTHRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHP 420  
QY 421 PHNQHCCHGHPGPPGHLRRRGKGRPFHCRQIGSVYRLPLRKGEVLPPLPEANFSPF 480  
Db 421 PHNQHCCHGHPGPPGHLRRRGKGRPFHCRQIGSVYRLPLRKGEVLPPLPEANFSPF 480  
QY 481 PLPHKHKLPLDNQNPQSVSSCPGKFKSGPPQVSMFTTTPFK 525  
Db 481 PLPHKHKLPLDNQNPQSVSSCPGKFKSGPPQVSMFTTTPFK 525  
RESULT 2  
A60488  
N:Altiternat names: autorsotet inhibition factor  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 19-Mar-1993 #sequence revision 23-Mar-1995 #text\_change 07-Jul-1995  
C:Accession: S35687; J2196; A60488  
R:Sorensen, C.B.; Krogn-Pedersen, H.; Petersen, T.E.  
FEBS Lett. 328, 285-290, 1993  
A:Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly  
A:Reference number: S35687; MUID:93351678  
A:Accession: S35687  
A:Molecule type: protein  
A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>  
A:Note: 355-Gln and 368-Tyr were also found  
R:Halkier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.  
Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
A:Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
A:Reference number: J2196; MUID:94220160  
A:Accession: J2196  
A:Molecule type: protein  
A:Residues: 1-23;35-54,'VK',57-101,'R','TVGEYTEG',116,'N',118,'R',120-136;137-145;150-20  
A:Experimental source: plasma  
R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.  
Thromb. Res. 60, 385-396, 1990  
A:Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
A:Reference number: A60488; MUID:91196010  
A:Accession: A60488  
A:Molecule type: protein  
A:Residues: 1-6,'X',8-15 <VES>  
C:Comment: This protein is a single-chained plasma protein which participates in transgl  
C:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: glycoprotein; plasma  
F:1-23/Domain: cystatin homology <CY1>  
F:122-207/Domain: cystatin homology (fragments) <CY2>  
F:7-424,60-71,87-108,165-346,180-203,258-288/Disulfide bonds: #status experimental

F:74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental  
Query Match 49.9%; Score 1514.5; DB 2; Length 445;  
Best Local Similarity 57.0%; Pred. No. 6e-103;  
Matches 293; Conservative 37; Mismatches 109; Indels 75; Gaps 9;  
QY 18 AVSPDTCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLDVENTTVYLVLDVQSD 77  
Db 1 AVNPDTCDAVEPEAVRALDLINK-GRDGYLFOLLRIADAHLDKVESIAVYLVLDKVESD 59  
QY 78 CSVLRSKYNWCEPPDSRRPGEIVIGQCKVIATRHSHESQDLRVIDFNCTSSVSALAN 137  
Db 60 CPVLSRKHWDCCLNVSRYPSETVIGQCKVITAITLAGPEDLVSNDFTCTSSVSALTN 119  
QY 138 TKDSPLVLDFFEDTERYRKQANKALEKYEENDDFASFRVDRIERVARVGGEGTYGV 197  
Db 120 IIDSPLVLDFFEDTLY-----ENSDPAPFRM-----RARGEGTSYFLD 159  
QY 198 FSVRNCPRHHPNPFVFCRADLFYDVEALDLESPKLNLYNCEVFDPOEHENINGVPP 257  
Db 160 FSVRNCSSHHFPRHSHIFGCRADLFYDVEASDLTPKDIVTNCVEFN-LKRRRESAVOH 218  
QY 258 HLGHFPFHGGHERSSTTKPKPKGSRDHHHPKPHGHPPPPDERHSHGP-----P 311  
Db 219 HLGRFPFHSGEHSHPAGPPPKPSGSKDHGHPHSHYFRCPPPLEKHKNSDSPPPQARAP 278  
QY 312 LPQGPPLPPLPMSCCSCQHATFTNGAQRHSHNNNSDLHPKHSHSQHHPHHPHHP 371  
Db 279 LPFPPLPGL-----RCPHPPFGTKGNHR-----PFDHSS 307  
QY 372 HEHDTHRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHP 431  
Db 308 DEH-----HNHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHP 352  
QY 432 GPPGHLRRRGKGRPFHCRQIGSVYRLPLRKGEVLPPLPEANFSPFPLPHKHKLPL 491  
Db 353 -PPRHSKERGPGKGFHFRWPTGYIHRPLSLKKGVLPLPEANFSPFLPHNHNLPQ 411  
QY 492 DNQPPQSVSSCPGKFKSGPPQVSMFTTTPFK 525  
Db 412 EIQAPQASCPGTFNFKFLHISKFFAYTLPK 445  
RESULT 3  
KGZHL  
histidine-rich glycoprotein precursor - plasmodium lophurae  
C:Species: Plasmodium lophurae  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 16-Jul-1999  
C:Accession: A22692  
R:Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.  
Nature 312, 616-620, 1984  
A:Title: Primary structure and genomic organization of the histidine-rich protein of  
A:Reference number: A22692; MUID:85061618  
A:Accession: A22692  
A:Molecule type: DNA  
A:Residues: 1-351 <RAV>  
A:Cross-references: GB:X01469; NID:g9997; PIDN:CAA25698.1; PID:g9999  
C:Comment: There are two copies of 16-residue repeats, two copies of 17-residue repea  
C:Genetics:  
A:Introns: 23/3  
C:Superfamily: plasmodium histidine-rich protein  
C:Keywords: glycoprotein; tandem repeat  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-47/Domain: propeptide #status predicted <PRO>  
F:48-351/Product: histidine-rich glycoprotein #status predicted <MAT>  
F:59-74,75-90/Region: 16-residue repeats  
F:91-107,108-123/Region: 17-residue repeats  
F:124-138,139-153/Region: 15-residue repeats  
F:173-301,312-331/Region: 10-residue repeats  
F:40/Binding site: carbohydrate (Asn) (covalent) #status predicted



Db 250 -----GRRHGHHG-----GGHGRSGSPGGRHGSGHHRRGRHGHHG-----R 290  
QY 308 HGPPPLPGPPPLPMSCCSQHATFTNGAQRH-SHNNNSDDLHPH-----KHHSH 357  
Db 291 HG-----SCSGSPCRHGHHGGHGGHGRSGSPGGRHGSGHHRRGRHG 337  
QY 358 EQH-----PHGHHHPH-----HHPEHDT-----RQHPHG--HHP-----HGHH 390  
Db 338 GRHGRSGSPGGRHGHHGGHGRHPPHPCGRHGRHGRSGHGRHGRHPPHPCGRHGRH 397  
QY 391 --PHGHH-----PHGHHPHGHH-----PH--CHDFQDY-GPCDP-----PPHNOGHC 427  
Db 398 GPPHHHHHDDRSRHHHHHHHHGCRPPPPHHGHHHFFPPFPWPPPPPPFPWPPHRRGGH 457  
QY 428 CHGH 431  
Db 458 CHHH 461  
RESULT 7  
KGHUH1  
N:kininogen, HMW precursor [validated] - human  
N:Alternate names: alpha-2-thiol proteinase inhibitor; preprokininogen; prokininogen  
N:Contains: bradykinin (kallidin I); HMW kininogen I; HMW kininogen II; low molecular weight kininogen  
C:Species: Homo sapiens (man)  
C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 08-Dec-2000  
A:Accession: A25276; S32422; A91153; A24871; A27899; A27699; A31905; A34030; S02482  
R:Ohkubo, I.; Kurachi, K.; Takasawa, T.; Shiokawa, H.; Sasaki, M.  
Biochemistry 23, 5691-5697, 1984  
A:Title: Isolation of a human cDNA for alpha-2-thiol proteinase inhibitor and its identification  
A:Reference number: A90490; MUID:85122621  
A:Accession: A01279  
A:Molecule type: mRNA  
A:Residues: 1-389 <OHK>  
A:Cross-references: GB:K02566; NID:q177889  
R:Takagaki, Y.; Kitamura, N.; Nakanishi, S.  
J. Biol. Chem. 260, 8601-8609, 1985  
A:Title: Cloning and sequence analysis of cDNAs for human high molecular weight and low molecular weight kininogen  
A:Reference number: A92544; MUID:85234582  
A:Accession: A25276  
A:Molecule type: mRNA  
A:Residues: 1-592, 'I', 594-644 <TAK>  
A:Cross-references: GB:M11437; NID:q186751; PIDN:AAB59550.1; PID:q385852  
R:Auerswald, E.A.; Roessler, D.; Mentele, R.; Assfalg-Machleidt, I.  
FEBS Lett. 321, 93-97, 1993  
A:Title: Cloning, expression and characterization of human kininogen domain 3.  
A:Reference number: S32422; MUID:93223854  
A:Accession: S32422  
A:Molecule type: mRNA  
A:Residues: 'ANSM', 253-377 <AUE>  
A:Note: differences are due to known cloning artifacts  
R:Lottspeich, F.; Kellermann, J.; Henschel, A.; Foerster, B.; Muller-Esterl, W.  
Eur. J. Biochem. 152, 307-314, 1985  
A:Title: The amino acid sequence of the light chain of human high-molecular-mass kininogen  
A:Reference number: A91153; MUID:86030270  
A:Accession: A91153  
A:Molecule type: protein  
A:Residues: 379-644 <LOT>  
A:Note: the bradykinin sequence preceding the light chain sequence was not determined in this clone  
R:Kellermann, J.; Lottspeich, F.; Henschel, A.; Mueller-Esterl, W.  
Eur. J. Biochem. 154, 471-478, 1986  
A:Title: Completion of the primary structure of human high-molecular-mass kininogen. The light chain sequence of the primary structure of human high-molecular-mass kininogen  
A:Reference number: A24871; MUID:86108361  
A:Accession: A24871  
A:Molecule type: protein  
A:Residues: 'Z', 20-380 <KEL1>  
R:Kellermann, J.; Lottspeich, F.; Henschel, A.; Mueller-Esterl, W.  
in Kinins IV, Greenbaum, L.M., and Margolis, H.S., ed., pp.85-89, Plenum Press, New York, 1985  
A:Title: Amino acid sequence of the light chain of human high molecular mass kininogen.  
A:Reference number: A27899  
A:Accession: A27899  
A:Molecule type: protein

A:Residues: 379-389, 'K', 390-407, 'Q', 409-644 <KEL2>  
R:Mindrou, T.; Carretero, O.A.; Proud, D.; Walz, D.; Scicli, A.G.  
Biochem. Biophys. Res. Commun. 152, 519-526, 1988  
A:Title: A new kinin moiety in human plasma kininogens.  
A:Reference number: A27699; MUID:88209021  
A:Accession: A27699  
A:Molecule type: protein  
A:Residues: 380-389 <MIN>  
R:Maeda, H.; Matsumura, Y.; Kato, H.  
J. Biol. Chem. 263, 16051-16054, 1988  
A:Title: Purification and identification of [hydroxyprolyl(3)]bradykinin in ascitic fluid  
A:Reference number: A31905; MUID:89034061  
A:Accession: A31905  
A:Molecule type: protein  
A:Residues: 381-389 <MAE>  
R:Sasaguri, M.; Ikeda, M.; Ideishi, M.; Arakawa, K.  
Biochem. Biophys. Res. Commun. 150, 511-516, 1988  
A:Title: Identification of [hydroxyproline(3)]-lysyl-bradykinin released from human plasma  
A:Reference number: A34030; MUID:88106632  
A:Accession: A34030  
A:Molecule type: protein  
A:Residues: 380-389 <SAS>  
R:Lenarcic, B.; Gabrijelcic, D.; Rozman, B.; Drobnic-Kosorok, M.; Turk, V.  
Biol. Chem. Hoppe-Seyler 369, 257-261, 1988  
A:Title: Human cathepsin B and cysteine proteinase inhibitors (CPis) in inflammatory diseases  
A:Reference number: S02482; MUID:89076517  
A:Accession: S02482  
A:Molecule type: protein  
A:Residues: 1-19; 189-192; 310-314; 381-389 <LEN1>  
R:Kato, H.; Matsumura, Y.; Maeda, H.  
FEBS Lett. 232, 252-254, 1988  
A:Title: Isolation and identification of hydroxyproline analogues of bradykinin in human plasma  
A:Reference number: A61495; MUID:88211869  
A:Accession: A61495  
A:Molecule type: protein  
A:Residues: 380-389 <KAT1>  
A:Experimental source: urine  
A:Note: this peptide had Pro-383 modified to 4-hydroxyproline  
A:Accession: B61495  
A:Molecule type: protein  
A:Residues: 381-389 <KAT2>  
A:Experimental source: urine  
A:Note: this peptide had Pro-383 modified to 4-hydroxyproline  
A:Accession: C61495  
A:Molecule type: protein  
A:Residues: 380-389 <KAT3>  
R:Lenarcic, B.; Krasovec, M.; Ritonja, A.; Olafsson, I.; Turk, V.  
FEBS Lett. 280, 211-215, 1991  
A:Title: Inactivation of human cystatin C and kininogen by human cathepsin D.  
A:Reference number: S14303; MUID:91192133  
A:Accession: S14303  
A:Molecule type: protein  
A:Residues: 264-359, 'N', 361-375 <LEN2>  
R:Little, S.S.; Johnson, D.A.  
Biochem. J. 307, 341-346, 1995  
A:Title: Human mast cell tryptase isoforms: separation and examination of substrate-sensitivity  
A:Reference number: S55239; MUID:95251593  
A:Accession: S55239  
A:Molecule type: protein  
A:Residues: 450-452, 'X', 454, 'X', 456 <LIT>  
R:Straczek, J.; Maschi, F.; le Nguyen, D.; Becchi, M.; Heulin, M.H.; Nabet, P.; Belle  
FEBS Lett. 373, 207-211, 1995  
A:Title: Purification from human plasma of a tetrapeptide that potentiates insulin-like growth factor release  
A:Reference number: S68059; MUID:96033974  
A:Accession: S68059  
A:Molecule type: protein  
A:Residues: 431-434 <STR>  
R:Kitamura, N.; Kitagawa, H.; Fukushima, D.; Takagaki, Y.; Miyata, T.; Nakanishi, S.  
J. Biol. Chem. 260, 8610-8617, 1985  
A:Title: Structural organization of the human kininogen gene and a model for its evolution  
A:Reference number: A92545; MUID:85234583  
A:Contents: annotation; gene organization  
R:Pierce, J.V.







Db 3 LIATSLVPDEHTPMSPVNTTTLKILORGKIMEIPPLYDPSQDDPEDG-----VNPV 57  
Qy 55 DAHLDRVENTTVYVLVDQESDCSVL---SRKYNDCEPPDSRSEIVIGOCKVIATRH 112  
Db 58 DPDLFTKNTNTEYDLVLKRPVDEAK---KIEVPASAPPKNKIVLYMYRTLK 114  
Qy 113 SHESODLKVDFNCTTSVSSALANTKDSPLVLIFFEDTERYKQANKALEKYKEENDF 172  
Db 115 ESE-----LIQLNAYTKRNLNLVKNNTI-----DREFDQACESLVKKLKKNDL 163  
Qy 173 ASERVD-----RIERVARVGGEGTGFV-----DFSVR 201  
Db 164 QNL-IDVVLSTGTYGTGITTPTLDGLQVHGRKPPVHYVYGLWRNEMTKNETRHVD 222  
Qy 202 NCPRHHPFRHNVFG-FCRADLFDVEALDLESPKNLVINCEVDPQSHENINGVPPHLG 260  
Db 223 HC-KHAFEMKSDMGEVCNYPHYEI-----VIGTMVQORDHNRDMPPH-- 268  
Qy 261 HPFWHGHE-----RSSTTKPPFKPH-----GSRDHHHPK-PHE 294  
Db 269 QRYHTPGRQDPVDDMSREIPPASIRPPPMNMHTRPQMPQOLPSVGFATFAPLPHQAPHN 328  
Qy 295 HGPPPPDPDRDHS-----GPPLOGPPPLLPMSCCSCQAHATFTNGAQRHSHNN 344  
Db 329 PGVSHPSYIAQTHYPLNMNFIQMPQMPPLP-----HOGYGMNGPSCSENN 379  
Qy 345 NSSDLHPKHHSHEQHGHGHHHPHAEHDTHTHQHGHGHHGHHGHHGHHGHH 404  
Db 380 N-PHQNHNYNDISHPN-HYSYDCGPNLYGFTPTYPDFHHPNQPH----- 424  
Qy 405 HPHCHDFQDYGCDPPPHNQHCCHGHPGPPGHLRRRGPGKGP--RP----- 449  
Db 425 -----QPQLSQNHTSQGSHOPGH-QGVNPDPPIISRPAALRTILLFIHD 470  
Qy 450 -FHCRIGSVYRLPLRKGEVLP 472  
Db 471 TFFCLRSSFLKFKVLEVNSITVL 494  
RESULT 12  
S22140  
nodulin Enod2 - Sesbania rostrata  
C:Species: Sesbania rostrata  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999  
C:Accession: S22140  
R:de Bruijn, F.J.  
submitted to the EMBL Data Library, November 1991  
A:Reference number: S22140  
A:Accession: S22140  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-330 <DEB>  
A:Cross-references: EMBL:X63339; NID:g21372; PIDN:CAA44939.1; PID:g21373  
C:Superfamily: proline-rich protein 3  
Query Match 6.5%; Score 197; DB 2; Length 330;  
Best Local Similarity 26.5%; Pred. No. 5.4e-07;  
Matches 71; Conservative 22; Mismatches 99; Indels 76; Gaps 14;  
Qy 246 POEHENINGVPPHGLHPHW-GGHERSSTTKPPKPGSGRDHHHP-----HKPHEHGPP-- 298  
Db 109 PPEYQ-----PPHENPPPEYQPPHE-----KPP--PEYQPPHEKPPPEYQPPHEKPPPEY 156  
Qy 299 PPDPERDHSHPPLPQGPPLLPMSCCSCQAHATFTNGAQRHSHNNSSDLHPKHHSHE 358  
Db 157 QPPHEKPPPEYQPPHEKPPPEYPPPEYKPPH-----EKPPSYE 194  
Qy 359 Q-HPHGHHPHAPHHEHDTHTHQHGHGHHGHHGHHGHHGHHGHHGHHGHHGHH 417  
Db 195 KPPPYEKPPEKPPPEYKPPHEKPPPEYKPPPEYKPPPEYKPPHEKPPPEYKPPPEY--- 251  
Qy 418 DPPPHNQHCCHGHPGPPGHLRRRGPGKGRPFHCRIGSVYRLPLRKGEVLPLEAFN 477

Db 252 -KPPHEK-----PPPEY---KPPHEKPP-----YEKPPHEK-----PPPEYK- 285  
Qy 478 PSFPLPHKHPLKPDNQPPQSVSESCLP 505  
Db 286 -----PPHEKPPPEYPPPYKPPPEYK 308  
RESULT 13  
S06733  
hydroxyproline-rich glycoprotein precursor - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jul-2000  
C:Accession: S06733  
R:Keller, B.; Lamb, C.J.  
Genes Dev. 3, 1639-1646, 1989  
A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein ge  
A:Reference number: S06733; MUID:90128263  
A:Accession: S06733  
A:Molecule type: DNA  
A:Residues: 1-620 <KEL>  
A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867  
C:Superfamily: hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein  
Query Match 6.5%; Score 197; DB 2; Length 620;  
Best Local Similarity 27.0%; Pred. No. 1.1e-06;  
Matches 75; Conservative 16; Mismatches 99; Indels 88; Gaps 15;  
Qy 273 TTKPPFKPHGSRDHHHPKHHEGPPPPPPDERDHSHPPLPQGPPLLPMSCCSCQAHATF 332  
Db 56 TTTTTPSRGHVSPRHAPPR---HAYPPP-----SHGHLPPSVGGPPP----- 94  
Qy 333 GTNGAQRHSHNNSSDLHPK-----HHSHEQHPHGHGHHHPHAEHDTHTHQHP---HGH 384  
Db 95 -----HRGHLPPSRGFNPPSPVISPSPSPSYGAPPSPHGHLPSPSPSPSHGH 148  
Qy 385 HPHGHHPHGH---PHGHHHPHGHHPHCHDFQDYGCDPPPHNQHCCHGHPGPPGHLRRRG 442  
Db 149 AP---PSGGHTPPRGQHPHPSHR-----RPSPPSR-----HGHPPPPPTTAQ--- 185  
Qy 443 PKGGRPPHCR-----QIGSVYRLPLRKGEVLPPLP--EAFNPSFPL-----PHKH--- 487  
Db 186---PPTTPIYSPQVQPPPTTSPPTTTHVQPTTSPSPSRGHQPPPTTHRHAPPTTHRHAPPT 243  
Qy 488 -----PLKPDNQPPQSVSESCLPQKFKSGFPQVS 516  
Db 244 HQPSPLRHLPPSRQPPPTTSPPPPAYAQS--PQPS 279  
RESULT 14  
Tl3049  
eyelid - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: Tl3049  
R:Treisman, J.E.; Lusk, A.; Rubin, G.M.; Heberlein, U.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z17592  
A:Accession: Tl3049  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2715 <TPE>  
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1  
C:Genetics:  
A:Gene: eld  
A:Cross-references: FlyBase:FBgn0003013  
C:Function:  
A:Description: could act as a transcription factor antagonistic to the wg pathway  
C:Keywords: DNA binding





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:03:16 ; Search time 13.48 Seconds  
(without alignments)  
1507.994 Million cell updates/sec

Title: DAVISN-P04196.PEP  
Perfect score: 3036  
Sequence: 1 MKALTAALLITLQYSCAVS.....GKFKSGFPQVSMFFTHTFPK 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3036	100.0	525	1	HRG_HUMAN
2	1725.5	56.8	526	1	HRG_RABIT
3	1327	43.7	396	1	HRG_BOVIN
4	306	10.1	351	1	HRPX_PLALO
5	235	7.7	378	1	FETB_RAT
6	226.5	7.5	388	1	FETB_MOUSE
7	226.5	7.5	661	1	KNG_MOUSE
8	221	7.3	382	1	FETB_HUMAN
9	217	7.1	644	1	KNG_HUMAN
10	211	6.9	449	1	CSUP_DROME
11	197	6.5	620	1	EXTN_TOBAC
12	193	6.4	621	1	KNH1_BOVIN
13	190	6.3	619	1	KNH2_BOVIN
14	186	6.1	309	1	NO75_SOYBN
15	186	6.1	476	1	KE4_MOUSE
16	179.5	5.9	657	1	KNOB_PLAFN
17	178.5	5.9	639	1	KNG_RAT
18	178	5.9	852	1	SRCH_RABIT
19	175.5	5.8	515	1	KE4L_CAEEL
20	174.5	5.7	251	1	PRP2_HUMAN
21	174	5.7	276	1	PRPL_HUMAN
22	174	5.7	434	1	NO75_LUPLU
23	173	5.7	549	1	DSX_DROME
24	171.5	5.6	247	1	PRP4_HUMAN
25	171	5.6	352	1	KE4_BRARE
26	170.5	5.6	331	1	PRP1_HUMAN
27	169	5.6	1185	1	DRPL_HUMAN
28	168.5	5.6	1183	1	DRPL_RAT
29	168	5.5	234	1	PRPM_HUMAN
30	168	5.5	395	1	SRP_MOUSE
31	166.5	5.5	424	1	S3BA_HUMAN
32	165.5	5.5	332	1	HRP1_PLAFA
33	163	5.4	428	1	FXB2_MOUSE

ALIGNMENTS

RESULT 1

ID	HRG_HUMAN	STANDARD	PRT	525 AA.
AC	P04196;			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich			
DE	glycoprotein) (HPRG).			
GN	HRG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived			
RT	from the nucleotide sequence of its cDNA.";			
RL	Biochemistry 25:2220-2225(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,			
RA	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;			
RT	"Evidence for the absence of intron H of the histidine-rich			
RT	glycoprotein (HRG) gene: genetic mapping and in situ localization of			
RT	HRG to chromosome 3q28-q29.";			
RL	Genomics 19:195-197(1994).			
RN	[4]			
RP	SEQUENCE OF 19-27.			
RC	TISSUE-Plasma;			
RX	MEDLINE=93092937; PubMed=1459097;			
RA	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,			
RA	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,			
RA	Hochstrasser D.F.;			
RT	"Plasma protein map: an update by microsequencing.";			
RL	Electrophoresis 13:707-714(1992).			
CC	-!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS			
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE			
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,			
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS			
CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN			
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD			
CC	COAGULATION CASCADE.			
CC	-!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS			
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-			
CC	RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.			
CC	-!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----

DR EMBL; M13149; AAA52694.1; -  
DR EMBL; AB005803; BAA21613.1; -  
DR EMBL; 217218; CAA78925.1; -  
DR PIR; A01287; KGHUGH.  
DR SWISS-2DPAGE; P04196; HUMAN.  
DR MIM; 142640; -  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
KW Glycoprotein; Heparin-binding; Repeat; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.  
FT DOMAIN 19 136 CYSTATIN-LIKE 1.  
FT DOMAIN 137 254 CYSTATIN-LIKE 2.  
FT DOMAIN 276 321 PRO-RICH.  
FT DOMAIN 350 497 PRO/HIS-RICH.  
FT DISULFID 24 504 BY SIMILARITY.  
FT DISULFID 78 89 BY SIMILARITY.  
FT DISULFID 105 126 BY SIMILARITY.  
FT DISULFID 203 417 BY SIMILARITY.  
FT DISULFID 218 241 BY SIMILARITY.  
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 100.0%; Score 3036; DB 1; Length 525;  
Best Local Similarity 100.0%; Pred. No. 1.7e-195;  
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKALIAALLIITQVSCAVPTDCSAVEPEAEKALDLINKRRDGYLFQLLRIADHLDR 60  
DB 1 MKALIAALLIITQVSCAVPTDCSAVEPEAEKALDLINKRRDGYLFQLLRIADHLDR 60  
QY 61 VENTTVYVLVDQESDCSVLSRKYWNCDEPPDSRPSSEIVIGQCKVIATRHSHESQDLR 120  
DB 61 VENTTVYVLVDQESDCSVLSRKYWNCDEPPDSRPSSEIVIGQCKVIATRHSHESQDLR 120  
QY 121 VIDFNCTTSVSSALANTKDSPLVIDFEDTERYRQANKALEKYEKENDDFASFRVDRI 180  
DB 121 VIDFNCTTSVSSALANTKDSPLVIDFEDTERYRQANKALEKYEKENDDFASFRVDRI 180  
QY 181 ERVARVGGEGTGYFVDFSVRNCPRHHPHNVGFCRADLFYDVEALDLESPKNLVIN 240  
DB 181 ERVARVGGEGTGYFVDFSVRNCPRHHPHNVGFCRADLFYDVEALDLESPKNLVIN 240  
QY 241 CEVFDPOEHENINGVPHLGHFPHWGHSRSTTKPPFKPHGRDHHHPKHEHGPPPP 300  
DB 241 CEVFDPOEHENINGVPHLGHFPHWGHSRSTTKPPFKPHGRDHHHPKHEHGPPPP 300  
QY 301 PDERDHSHGPPPLPQGGPPPLPMSCCSQHATGTGNAQRSHNNNSDDLPHKHHSHSQH 360  
DB 301 PDERDHSHGPPPLPQGGPPPLPMSCCSQHATGTGNAQRSHNNNSDDLPHKHHSHSQH 360  
QY 361 PPHGHHPAHHPHEDHROHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPH 420  
DB 361 PPHGHHPAHHPHEDHROHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPH 420  
QY 421 PHNQGHCHGHGPPPPGHLRRRPGKGRPFHCRQIGSVYRLPLRKEVLPPLPEANFPSPF 480  
DB 421 PHNQGHCHGHGPPPPGHLRRRPGKGRPFHCRQIGSVYRLPLRKEVLPPLPEANFPSPF 480  
QY 481 PLPHHKLPLKPDQPPQSVSESCPGKFKSGFPQVSMFTHTFPK 525

DB 481 PLPHHKLPLKPDQPPQSVSESCPGKFKSGFPQVSMFTHTFPK 525

RESULT 2  
HRG\_RABIT  
ID HRG\_RABIT STANDARD; PRT; 526 AA.  
AC Q28640;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Histidine-rich glycoprotein precursor (Histidine-proline rich  
DE glycoprotein) (HPRG) (Fragment).  
GN HRG.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.  
RC TISSUE=Serum;  
RX MEDLINE=96229917; PubMed=8639676;  
RA Borza D.-B., Tatum F.M., Morgan W.T.;  
RT "domain structure and conformation of histidine-proline-rich  
RT glycoprotein.";  
RL Biochemistry 35:1925-1934(1996).  
CC -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS  
CC HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE  
CC FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,  
CC AND THE LYSINE-BINDING SITE OF PLASMINOGEN ON THE BASIS OF ITS  
CC HOMOLGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN  
CC MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD  
CC COAGULATION CASCADE.  
CC -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS  
CC PROTEIN HAS MANY INTERNAL REPEATS. 15 TANDEM REPEATS OF A 5-  
CC RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH  
CC REGION.  
CC -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----

DR EMBL; U32189; AAC48516.1; -  
DR InterPro; IPR000010; Cystatin.  
DR Pfam; PF00031; cystatin; 1.  
DR SMART; SM00043; CY; 2.  
KW Glycoprotein; Heparin-binding; Repeat; Signal.  
FT NON\_TER 1 1  
FT SIGNAL <1 8 POTENTIAL.  
FT CHAIN 9 526 HISTIDINE-RICH GLYCOPROTEIN.  
FT DOMAIN 9 126 CYSTATIN-LIKE 1.  
FT DOMAIN 127 243 CYSTATIN-LIKE 2.  
FT DOMAIN 251 296 PRO-RICH.  
FT DOMAIN 329 498 PRO/HIS-RICH.  
FT DISULFID 14 505 BY SIMILARITY.  
FT DISULFID 68 79 BY SIMILARITY.  
FT DISULFID 95 116 BY SIMILARITY.  
FT DISULFID 193 415 BY SIMILARITY.  
FT DISULFID 207 230 BY SIMILARITY.  
FT DISULFID 272 302 POTENTIAL.  
FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 303 304 CLEAVAGE (BY PLASMIN).  
FT SITE 421 422 CLEAVAGE (BY PLASMIN).  
SQ SEQUENCE 526 AA; 58877 MW; 810F23D3D67D93D42 CRC64;

Query Match		56.8%;	Score 1725.5;	DB 1;	Length 526;
Best Local Similarity		60.6%;	Pred. No. 3.9e-108;		
Matches		328;	Conservative	48;	Mismatches 122; Indels 43; Gaps 7;
Qy	12	TLQYSCAVPTDCSAVEPEAKALDLINKRRRDGYLFOLLRIADAHLDRENTVTVYLV	71		
Db	2	TLQCSWALTPTDCKTKPLAEKALDLINKRRRDGYLFOLLRVADAHLDGAESATVYLV	61		
Qy	72	DVOESDCSVLSRKYNDCPEPDSRRPSEIVIGOCKVIATRIHSHESODLRVIDFNCTTSSV	131		
Db	62	DKKETDCSVLSRKHWEDCDPLTKRYSDVIGQCKVIATRIHSHESODLRVIDFNCTTSSV	121		
Qy	132	SSALANTKDSVLDIFFEDTERYKQANKALEKYEENDDFASFRVDRIERVARVRCGEG	191		
Db	122	SSALANTKDSVLDIFEDTEPRKRSADKALEVYKSESEAYASFRVDRVERTRVKGGER	181		
Qy	192	TGYFVDSVRNCPHRRHPRHNVFGCRADLFYDVEALDLSPKNLVINCEVDFDQEHEN	251		
Db	182	TNYVYDFSVRNCSRHFRHP-AFGFCRADLSFDVEASNLLENPEVDIISCEVFNPEEHN	240		
Qy	252	INGVPPHGLGHFPHGGHRSSTTKPPFKPHGSRDHHHPKHKEHCGPPPPDEROHSHP	311		
Db	241	ISGRFPHLG-----KTPLTGDSRDHHHPKHKEHCGPPPPQEGEDFSEGPP	286		
Qy	312	LPQGPPLLPMSKSCSQHATFTNGAQRHSHNNSSDL-----HPKHHSHEQHGHHPH	367		
Db	287	LOGGTPPLSPFPCRCRHRPCTNETHRPHRISVNIHRPPHGHHPHGHHPH	346		
Qy	368	A-----HH-----PHEHDTQRHDPHGHHPHGHHPHGHHPHGHHPHGHHPH	408		
Db	347	GPPPHGHPHGHPPHPPHGHPPHGHPPHGHPPHGHPPHGHPPHGHPPHGHPPH	406		
Qy	409	HDFOYDGCDDPPHNOG----HCCHGHPGPPHGLRRRRGKGRPFCHRCROIGSVYRLPPL	464		
Db	407	HGFHDHGCDDPSHKEGQDLH-QUAMGPPPKHKGKRGKGFHRRIGSVYQLPPL	465		
Qy	465	RKGEVLPLPEANFPSPFLPHHKKHLKPDNQPPQSVSCFCGKFKSGFPQVSMFTHTFP	524		
Db	466	QKGEVLPLPEANFPOLLNRNTHPLKPEIQPPQVASERCPEFNGEFAQLSKFPSTFP	525		
Qy	525	K 525			
Db	526	K 526			
RESULT 3					
HRG_BOVIN					
ID	HRG_BOVIN	STANDARD;	PRT;	396	AA.
AC	P33433;				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)				
DE	(HPRG) (fragments).				
GN	HRG.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE, AND DISULFIDE BONDS.				
RX	MEDLINE=93351678; PubMed=8348977;				
RA	Soerensen C.B., Krogh-Pedersen H., Petersen T.E.;				
RT	"Determination of the disulphide bridge arrangement of bovine				
RT	histidine-rich glycoprotein."				
RL	FEBS Lett. 328:285-290(1993).				
CC	-!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS				
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE				
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,				
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS				

CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN				
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD				
CC	COAGULATION CASCADE.				
CC	-!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS				
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-				
CC	RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.				
CC	-!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.				
DR	PIR: S35687; S35687.				
DR	pfam: PF00031; cystatin; 1.				
KW	Glycoprotein; Heparin-binding; Repeat.				
FT	DOMAIN 1 102				
FT	DOMAIN 103 169				
FT	DOMAIN 191 238				
FT	DOMAIN 243 368				
FT	DISULFID 7 375				
FT	DISULFID 56 67				
FT	DISULFID 77 92				
FT	DISULFID 123 297				
FT	DISULFID 137 160				
FT	DISULFID 212 242				
FT	NON_CONS 52 53				
FT	CARBOHYD 70 70				
FT	NON_CONS 71 72				
FT	NON_CONS 78 79				
FT	CARBOHYD 91 91				
FT	NON_CONS 103 104				
FT	CARBOHYD 122 122				
FT	NON_CONS 163 164				
FT	CARBOHYD 220 220				
FT	NON_CONS 263 264				
FT	NON_CONS 303 304				
FT	VARIANT 86 86				
FT	VARIANT 309 309				
FT	VARIANT 322 322				
FT	VARIANT 396 AA; 44470 MW; 128A8223499DE6FC CRC64;				
SQ					
Query Match					
Best Local Similarity 43.7%; Score 1327; DB 1; Length 396;					
Matches 262; Conservative 38; Mismatches 90; Indels 124; Gaps					
Qy	18	AVSPDTCSAVEPEAEKALDLINKRRRDGYLFQLLRITADAHLDRENTVTVYLVLDVOESD	77		
Db	1	AVNPTGCDAVEPAVAVRALDLINK-GRDGYLFQLLRVADAHLDKVESIAVYLV-----ESD	55		
Qy	78	CSVLRSKYWDCPEPDSRRPSEIVIGOCKVIATPHSHESODLRVIDFNCTTSSVSSALAN	137		
Db	56	CPVLSRKHWDCDCE-----LNVTVIGOCKLAG-----PEDLSVNDFNCTTSSVSSALTN	103		
Qy	138	TKDSPVLIDFFEDTERYRKQANKALEKYEENDDFASFRVDRIERVARVRCGEGTYFYVD	197		
Db	104	-----MRARGEGTSFYFLD	117		
Qy	198	FSVRNCPRHHPHPRHNVFGCRADLFYDVEALDLSPKNLVINCEVDFDQEHENINGVPP	257		
Db	118	FSVRNCSHHFPRH-HIFGFCRADLFYDVEASDLTPDKIVTNCEVF-----HRRFSAVQH	172		
Qy	258	HLGHPFHGWGHERSTTKPFKPHGSRDHHHPKHKEHCGPPPPDERDHSHP-----P	311		
Db	173	HLGRPFHSGEHEHSPAGRPFPKPSGKDGHGHPHESYNFCPPPLEKHNSDPSFPQARAP	232		
Qy	312	LPQGPPLLPMSKSCSQHATFTNGAQRHSHNNSSDLPHKHSHEQHGHHPHHPHHPH	371		
Db	233	LPFPPLGL-----RCPHPFGTKG-----NHRP-----P	256		
Qy	372	HEHDTQRHQPHGHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPH	431		
Db	257	HDHSSDE-----HHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH	303		
Qy	432	GPPPGHLRRGPGKGRPFHRCQIGSVYRLPLRKGEVLPLPEANFPSPFLPHKHPKLP	491		
Db	304	PPRHSKERGPGKGRFHRWPTGYIHLRPLSLKGEVLPLPEANFPSPFLPNNHPNQ	362		





```
Db 120 KRAM-----FHVKNRRLVLPAYNCTLRVSKRKHSMCPDPCPHVPL--SAPSVLEAAT 173
Qy 160 KALEYKKEEN--DDFASRVRDIARVARGEGTGYFVDFSVR-----NCPRHIF 208
Db 174 ESLAKFNENPSQALVAVTK-----ATTQWVGSFYVEYLKESPTQSDSCSLOAS 229
Qy 209 PRHPNVFGCRADLFYDVEALDLESP-----KNLVINCEVDFDPQEHENINGVPPHILG 260
Db 230 DSEP--VGLCOGSL-----IKSPGVPPQRFKKTIVVSCFEFSQDQ-----VP---- 270
Qy 261 HPFHGCHGERST---TKPPFK---PHGSRDHHHPKHEHGP---PPPPDERDHDHGHGPPPL 312
Db 271 -----GGENPADTQAKLPQKNTAPTSPSTAPRGSIQHLPEQEEPEDESKSPPEPF 325
Qy 313 P 313
Db 326 P 326

RESULT 6
FETB_MOUSE
ID FETB_MOUSE STANDARD; PRT; 388 AA.
AC Q90XCL1
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Fetuin-B precursor (IHL685).
GN FETUB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=Liver;
RX MEDLINE=20407138; PubMed=10947975;
RA Olivier E., Soury E., Ruminy P., Husson A., Parmentier F., Daveau M.,
RA Salier J.-P.;
RT "Fetuin-B, a second member of the fetuin family in mammals.";
RL Biochem. J. 350:589-597(2000).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- TISSUE SPECIFICITY: LIVER, LUNG AND TONGUE.
CC -!- SIMILARITY: BELONGS TO THE FETUIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ242927; CAB62541.1; -
DR MGD; MGI:1890221; Fetub.
DR InterPro; IPR00010; Cystatin.
DR InterPro; IPR001363; Fetuin.
DR Pfam; PF00031; cystatin; 2.
DR SMART; SM00043; Cy; 2.
DR PROSITE; PS01254; FETUIN_1; 1.
DR PROSITE; PS01255; FETUIN_2; 1.
KW Glycoprotein; signal; Repeat.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 388 FETUIN-B.
FT DOMAIN 27 152 CYSTATIN-LIKE 1.
FT DOMAIN 153 276 CYSTATIN-LIKE 2.
FT DISULFID 96 107 BY SIMILARITY.
FT DISULFID 120 140 BY SIMILARITY.
FT DISULFID 154 157 BY SIMILARITY.
FT DISULFID 217 224 BY SIMILARITY.
FT DISULFID 237 263 BY SIMILARITY.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).
```

```
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 388 AA; 42712 MW; 39CFAD76A9D8DC2B CRC64;

Query Match 7.5%; Score 226.5; DB 1; Length 388;
Best Local Similarity 26.1%; Pred. No. 2.4e-08;
Matches 91; Conservative 52; Mismatches 135; Indels 71; Gaps 17;

Qy 4 LIAALLLITLQYSC-----AVSPDTC--SAVEPEAEKALDLINKRRRDG 45
Db 3 LRLVLVLTACACMARSPAPPLPQRPPLPLHPLGNCNDSEVLAVAGFALQINRDKDG 62
Qy 46 YLFQLRIADAHLDREVNT--TVYLLVDYQESDCSVLSRKYWNDCPEPPSRPSEIVGQ 104
Db 63 YMLSLNRVHDVREHYQEDMGSLFYLTLDVLETDCHVLSRKAQDKCP---RMFYESVYQG 119
Qy 105 CKVIATRHSHESQDLRVI---DFNCTTSSVSALANT--KDSPLVLDFFEDTERYRKQAN 159
Db 120 KRAM-----FHINKRRLVLPAYNCTLRVSKRKHHTTCTDPCPSPIDLSNPSA--LEAAT 173
Qy 160 KALEYKKEENDDFASRVRDIARVARGEGTGYFVDFSVR-----NCPRHIFPR 210
Db 174 ESLAKFNENPSQALVAVTK---AMNQWVSGPAYVEYLKESPTQSDSCSLOHSDS 231
Qy 211 HPNVFGCRADLFY-DVEALDLESP---KNLVINCEVDFDPQEHENINGVPPHILGHPHFGG 267
Db 232 EP--VGICQGSTVQSSLRHVPLIQPVKSVTVTCPEFESQAO-----VP-----G 274
Qy 268 HERSSTTKPPFK-----PHGSRDHHHPKHEHGP-----PPPPDERDHDHSHG 309
Db 275 DENPAVTGQPKLPQKNTAPTSPSTAPRGSIQHLPELDDEKPEESKG 323

RESULT 7
KNG_MOUSE
ID KNG_MOUSE STANDARD; PRT; 661 AA.
AC O08677; O08676;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kininogen precursor [Contains: Bradykinin].
GN KNG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS HMW AND LMW).
RC STRAIN=C57BL/6 x CBA; TISSUE=Liver;
RA Takano M., Kondoh J., Yamada K., Okamoto H.;
RT "Molecular cloning of cDNAs for mouse low- and high- molecular
RT kininogen.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: (1) KININOGENS ARE INHIBITORS OF THIOLESTERASES; (2)
CC HMW-KININOGEN PLAYS AN IMPORTANT ROLE IN BLOOD COAGULATION BY
CC HELPING TO POSITION OPTIMALLY PREKALLIKREIN AND FACTOR XI NEXT TO
CC FACTOR XII; (3) HMW-KININOGEN INHIBITS THE THROMBIN- AND PLASMIN-
CC INDUCED AGGREGATION OF THROMBOCYTES; (4) THE ACTIVE PEPTIDE
CC BRADYKININ THAT IS RELEASED FROM HMW-KININOGEN SHOWS A VARIETY OF
CC PHYSIOLOGICAL EFFECTS: (4A) INFLUENCE IN SMOOTH MUSCLE
CC CONTRACTION, (4B) INDUCTION OF HYPOTENSION, (4C) NUTRIENT AND
CC DIURESIS, (4D) DECREASE IN BLOOD GLUCOSE LEVEL, (4E) IT IS A
CC MEDIATOR OF INFLAMMATION AND CAUSES (4E1) INCREASE IN VASCULAR
CC PERMEABILITY, (4E2) STIMULATION OF NOCICEPTORS (4E3) RELEASE OF
CC OTHER MEDIATORS OF INFLAMMATION (E.G. PROSTAGLANDINS), (4F) IT HAS
CC A CARDIOPROTECTIVE EFFECT (DIRECTLY VIA BRADYKININ ACTION,
CC INDIRECTLY VIA ENDOTHELIUM-DERIVED RELAXING FACTOR ACTION); (5)
CC LMW-KININOGEN INHIBITS THE AGGREGATION OF THROMBOCYTES; (6) LMW-
CC KININOGEN IS IN CONTRAST TO HMW-KININOGEN NOT INVOLVED IN BLOOD
CC CLOTTING (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; HMW (SHOWN HERE) AND LMW; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
```





















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:02:36 ; Search time 30.96 Seconds  
(without alignments)  
2933.538 Million cell updates/sec

Title: DAVISN-P04196.PEP  
Perfect score: 3036  
Sequence: 1 MKALIAALLITLQYSCAVS.....GKFKSGFPQVSMFFTHTPPK 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTEMBL19.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1901.5	62.6	546	6 Q9BGU1	Q9bgul Bos taurus
2	1750	57.6	525	11 Q99PS6	Q99ps6 mus musculus
3	1746	57.5	525	11 Q9ESB3	Q9esb3 mus musculus
4	1742	57.4	525	11 Q99PS5	Q99ps5 mus musculus
5	1733	57.1	515	11 Q99PS7	Q99ps7 rattus norv
6	1725	56.8	525	11 Q99PS8	Q99ps8 rattus norv
7	1705.5	56.2	510	11 Q9ESB2	Q9esb2 rattus norv
8	667	22.0	214	6 Q9TS85	Q9ts85 bos taurus
9	269	8.9	735	5 Q9NES7	Q9nes7 caenorhabdi
10	266	8.8	140	5 Q26056	Q26056 plasmodium
11	256	8.4	594	5 Q9VEP4	Q9vep4 drosophila
12	238.5	7.9	294	2 Q9KI87	Q9ki87 bacillus ce
13	232	7.6	623	5 P90533	P90533 dictyosteli
14	229.5	7.6	296	2 Q9KI89	Q9ki89 bacillus ce
15	223.5	7.4	471	5 Q9UAY0	Q9uay0 caenorhabdi
16	222.5	7.3	372	5 Q9VSZ8	Q9vsz8 drosophila

17	217	7.1	485	5	076876	076876 drosophila
18	216.5	7.1	273	2	Q9KI88	Q9ki88 bacillus ce
19	213.5	7.0	102	5	Q9VUE1	Q9vuel drosophila
20	213	7.0	774	3	042918	042918 schizosacch
21	209.5	6.9	1002	3	013849	013849 schizosacch
22	208	6.9	1033	10	Q9XEK5	Q9xek5 chlamydomon
23	203	6.7	403	16	Q98H88	Q98h88 rhizobium l
24	202.5	6.7	940	4	Q96QC0	Q96qc0 homo sapien
25	202.5	6.7	1571	11	054978	054978 mus musculu
26	201.5	6.6	130	2	Q9KI77	Q9ki77 bacillus an
27	201.5	6.6	133	2	Q9KI76	Q9ki76 bacillus an
28	201.5	6.6	265	2	Q9KI91	Q9ki91 bacillus an
29	201	6.6	112	11	Q9GR74	Q9gr74 mus musculu
30	200.5	6.6	356	13	073679	073679 brachydanio
31	198.5	6.5	356	13	Q98UK5	Q98uk5 brachydanio
32	198.5	6.5	796	5	017532	017532 caenorhabdi
33	197	6.5	330	10	041402	041402 sesbania ro
34	197	6.5	341	5	Q9VV44	Q9vv44 drosophila
35	197	6.5	815	5	Q9V9W8	Q9v9w8 drosophila
36	197	6.5	865	5	Q9U7D4	Q9u7d4 neospora ca
37	197	6.5	1480	10	Q9LIE8	Q9lie8 arabidopsis
38	196.5	6.5	780	4	Q96P76	Q96p76 homo sapien
39	195	6.4	554	5	Q9W4C1	Q9w4c1 drosophila
40	194.5	6.4	2715	5	061603	061603 drosophila
41	194	6.4	884	4	000302	000302 homo sapien
42	194	6.4	2703	5	Q9VEG7	Q9veg7 drosophila
43	193.5	6.4	350	11	Q63179	Q63179 rattus norv
44	193	6.4	209	11	Q99KB0	Q99kb0 mus musculu
45	193	6.4	778	5	Q965D9	Q965d9 drosophila

ALIGNMENTS

RESULT 1						
Q9BGU1						
ID	Q9BGU1	PRELIMINARY;	PRT;	546	AA.	
AC	Q9BGU1;					
DT	01-JUN-2001 (TrEMBLrel. 17, Created)					
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)					
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)					
DE	HISTIDINE-RICH GLYCOPROTEIN.					
GN	BTHRG.					
OS	Bos taurus (Bovine).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
OC	Bovidae; Bovinae; Bos.					
OX	NCBI_TaxID=9913;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,					
RA	Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;					
RT	"Molecular diversity of mammalian histidine-rich glycoprotein."					
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AB055894; BAB33091.1;					
DR	InterPro; IPR000010; Cystatin.					
DR	Pfam; PF000031; cystatin; 1.					
DR	SMART; SM00043; CY; 2.					
SQ	SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;					

Query Match	62.6%;	Score	1901.5;	DB 6;	Length	546;
Best Local Similarity	63.2%;	Pred. No.	2.8e-156;			
Matches	349;	Conservative	45;	Mismatches	125;	Indels 33; Gaps 5;
Qy	1	MKALIAAL---	LLLTLYQSCAVSPTDCS	AVEPEAKDLINKRRRGGYLFQLLR	ADAH	57
Db	1	MKVITALLS	LLLTITQQCS	CAVNPFGD	AVPFAVRALDLINKRWGDLFQLLR	RVADAH 60
Qy	58	LDRVENTVYV	VLVDVQES	CVLSRKYNDCPEPPDS	RRPSEIVIGQCKVIATRHS	ESQ 117
Db	61	LDRVESTAVYV	VLVDVRES	DCPVLRSKRHWDC	ELNVSRPSETVIGQCKVIAT	LLAGPE 120



```
Db 179 ERVIRARGERTNTYVEFSMRNCSTQHPRSPVLFVGFRCALLSYSIETSDLETPDSIDIN 238
Qy 241 CEVDPDOEHENINGVPPHGLHPFHGHESSSTTKPKPKHGSRDHHPHHPKHEHGPPPP 300
Db 239 CEVFNIEDHKDTSMDK-----HW-GHERPLCDKHLCKLSGSRDHHHTHTKDLGCGPPP 291
Qy 301 PDERDHSHPPLPOGPPPLPMSSCSQHATFGTNGAQRHSHNN--NSSDLH--PHKHH 355
Db 292 PEGKNSDRPRLQEGALPQLPGYP--PHS--GANRTHRPSYNSHCNEHPCGHPRPHGH 347
Qy 356 SHEQHP-----HGHPHAPHHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGHCHD 410
Db 348 PSHHPHPGHSHGHHPHSHSHSGHHPHHPHGHHPHGHHPHGHHPHGHHPHGHCHD 407
Qy 411 FQDYGCPDPPHNO---GHCCHGHPGPPGHLRRRPGKPPGPPHCRQIGSVYRLPPLRKG 467
Db 408 FLDYGCPCDPSNSQELKQYHRGYPGPHGHSKRKPGKGLFPFHQQIGYVYRLPPLNIG 467
Qy 468 EVLPLPEANFPFPLPHHKKHLKPDNQPPQSVSESCPGKFKGFPQVSMFFTHFTPK 525
Db 468 EVLTLPEANFPFSLPNCNRSLOEIQPFQPTASRSCPGKFESEFPQISKFFGYTPPK 525
```

## RESULT 4

```
Q99PS5 ID Q99PS5 PRELIMINARY; PRT; 525 AA.
AC Q99PS5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).
GN MMHRG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAUSBERG R.;
RA Tsuchida N., Wakabayashi S., Jahnen-Dechent W., Koide T.;
RT "Structure of mouse histidine-rich glycoprotein gene.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055898; BAB33095.1; -
DR EMBL; BC011168; AAH11168.1; -
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;
```

```
Query Match 57.4%; Score 1742; DB 11; Length 525;
Best Local Similarity 62.1%; Pred. No. 1.7e-142;
Matches 334; Conservative 51; Mismatches 127; Indels 26; Gaps 9;

Qy 1 MKALIAALLLITLOYSACVSTDCSAVEPEAEKALDLINKRRRGYLFQLLRIADHLDR 60
Db 1 MKVLTALLTLTQCSHALSPNCDSKPLAEKVLIDLINKRRRGYTFQLLRVSDAHLDR 60
Qy 61 VENTTVYVLVDVQESDCSVLSRYWNCDEPPDSRRPSEIVIGOCKVIAIRHSHESODLR 120
Db 61 AGTATVYVLDVIESDCWVLSLTKAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLS 118
Qy 121 VIDNCTTSSVSSALANTKDSPLIDFEDTERYRKANKALEYKKEENDDFASFVRDRI 180
Db 119 VNGVYVLDVIESDCWVLSLTKAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLS 178
Qy 241 ERVARVREGGEGTGTFVDFSVNRCPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 300
Db 181 ERVARVREGGEGTGTFVDFSVNRCPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 240
```

```
Db 179 ERVIRARGERTNTYVEFSMRNCSTQHPRSPVLFVGFRCALLSYSIETSDLETPDSIDIN 238
Qy 241 CEVDPDOEHENINGVPPHGLHPFHGHESSSTTKPKPKHGSRDHHPHHPKHEHGPPPP 300
Db 239 CEVFNIEDHKDTSMDK-----HW-GHERPLCDKHLCKLSGSRDHHHTHTKDLGCGPPP 291
Qy 301 PDERDHSHPPLPOGPPPLPMSSCSQHATFGTNGAQRHSHNN--NSSDLH--PHKHH 355
Db 292 PEGKNSDRPRLQEGALPQLPGYP--PHS--GANRTHRPSYNSHCNEHPCGHPRPHGH 347
Qy 356 SHEQHP-----HGHPHAPHHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGHCHD 410
Db 348 PSHHPHPGHSHGHHPHSHSHSGHHPHHPHGHHPHGHHPHGHHPHGHHPHGHCHD 407
Qy 411 FQDYGCPDPPHNO---GHCCHGHPGPPGHLRRRPGKPPGPPHCRQIGSVYRLPPLRKG 467
Db 408 FLDYGCPCDPSNSQELKQYHRGYPGPHGHSKRKPGKGLFPFHQQIGYVYRLPPLNIG 467
Qy 468 EVLPLPEANFPFPLPHHKKHLKPDNQPPQSVSESCPGKFKGFPQVSMFFTHFTPK 525
Db 468 EVLTLPEANFPFSLPNCNRSLOEIQPFQPTASRSCPGKFESEFPQISKFFGYTPPK 525
```

## RESULT 5

```
Q99PS7 ID Q99PS7 PRELIMINARY; PRT; 515 AA.
AC Q99PS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN 2.
GN RNHRG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055896; BAB33093.1; -
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;
```

```
Query Match 57.1%; Score 1733; DB 11; Length 515;
Best Local Similarity 62.2%; Pred. No. 1e-141;
Matches 332; Conservative 48; Mismatches 126; Indels 28; Gaps 7;
```

```
Qy 1 MKALIAALLLITLOYSACVSTDCSAVEPEAEKALDLINKRRRGYLFQLLRIADHLDR 60
Db 1 MKVLTALLTLTQCSHALSPNCDSKPLAEKVLIDLINKRRRGYTFQLLRVSDAHLDR 60
Qy 61 VENTTVYVLVDVQESDCSVLSRYWNCDEPPDSRRPSEIVIGOCKVIAIRHSHESODLR 120
Db 61 VETATVYVLDVIESDCWVLSLTKAQDECLP--ARMTSEIVIGOCKVIAIRHSHESODLS 118
Qy 121 VIDNCTTSSVSSALANTKDSPLIDFEDTERYRKANKALEYKKEENDDFASFVRDRI 180
Db 119 VNGVYVLDVIESDCWVLSLTKAQDECLP--ARMTSEIVIGOCKVIAIRHSHESODLS 178
Qy 181 ERVARVREGGEGTGTFVDFSVNRCPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 240
Db 179 ERVIRGREGGERTSYFIEFSVRNCSTQHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 238
Qy 241 CEVDPDOEHENINGVPPHGLHPFHGHESSSTTKPKPKHGSRDHHPHHPKHEHGPPPP 300
Db 239 CEVNTEDLNRSMDKPH-----RGHEHPHCDKHLCKLSGSRDHHHTHTKTHEIGCPPP 291
```



```
QY 415 GPCDPPPHQHCCHGHPGPPGHLRRRGGKGRPFHQICQISVYRLPLRKGEVLPLPE 474
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 GPCDPPPSQYH--QCHGPPHSGSRKRGKGLFFPHQRIQIGVYVRLPLNVEVLTPPE 459

QY 475 ANFPFPLPHHKLKPDNQPPQSVSESCPKGKSGGPOVSMFETHFPK 525
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 ANFPFSLPNCNRPQPEIRPFQASKSCPKGKFGKFPQVSNFETHPTPK 510

RESULT 8
Q9TS85 PRELIMINARY; PRT; 214 AA.
AC Q9TS85:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE HISTIDINE-RICH GLYCOPROTEIN-FACTOR XIIIA SUBSTRATE (FRAGMENTS).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=94220160; PubMed=7909439;
RA Haikier T., Andersen H., Vestergaard A., Magnusson S.;
RT "Bovine histidine-rich glycoprotein is a substrate for bovine plasma
RT factor XIIIA."
RL Biochem. Biophys. Res. Commun. 200:78-82(1994).
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
FT NON_TER 1
FT NON_CONS 23 24
FT NON_CONS 91 92
FT NON_CONS 120 121
FT NON_CONS 180 181
FT NON_CONS 198 199
FT NON_TER 214
SQ SEQUENCE 214 AA; 23982 MW; B8989492D6097A35 CRC64;

Query Match 22.0%; Score 667; DB 6; Length 214;
Best Local Similarity 52.0%; Pred. No. 6.5e-50;
Matches 142; Conservative 21; Mismatches 48; Indels 62; Gaps 7;

QY 18 AVSPDTCASVPEAEKALDLINKRRRDGYLFQLLRIDAHLDRVENTVYVLVDVQESD 77
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 AVNPTGDAVEPVAVRALDLINK-----VADAHLDKVESIAVYVYLDVVKESD 48

QY 78 CSVLSRKYWNCDPPDRPSEIVIGQCKVIATRHSHESQDLRVDFNCTTSSVSSALAN 137
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49 CPVLSRKHWDDCELNVSRYPSTVIGQCKVIAITLLAGPEDLR-----TVGEYTEGAN 101

QY 138 TK--DSPVLIDFFEDTEREYRKQANKALEKYRENDDFASFRVDRTERVARVGGEGTYF 195
|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 TRIDSPVLFDFEDTTL-----GGEGTSYF 128

QY 196 VDFSVRNCPRIHFRHPNVPFGCRADLFYDVEALDPSKPLVINCEVDFDQEHENINGV 255
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 LDFSVRNCSHHFPRSHIFGCRADLFYDVEASDLPTPKDIVTNCFVN---LKNFSAV 185

QY 256 PPLHGHPHWGSGHRSSTTKPPFKPHGSRDHHH 288
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 QHHLGRPFHWGGH---GNHRPPH--DNSSDEHH 213

RESULT 9
Q9NES7 PRELIMINARY; PRT; 735 AA.
AC Q9NES7:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HISTIDINE-RICH GLYCOPROTEIN-FACTOR XIIIA SUBSTRATE (FRAGMENTS).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=94220160; PubMed=7909439;
RA Haikier T., Andersen H., Vestergaard A., Magnusson S.;
RT "Bovine histidine-rich glycoprotein is a substrate for bovine plasma
RT factor XIIIA."
RL Biochem. Biophys. Res. Commun. 200:78-82(1994).
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
FT NON_TER 1
FT NON_CONS 23 24
FT NON_CONS 91 92
FT NON_CONS 120 121
FT NON_CONS 180 181
FT NON_CONS 198 199
FT NON_TER 214
SQ SEQUENCE 214 AA; 23982 MW; B8989492D6097A35 CRC64;

Query Match 8.9%; Score 269; DB 5; Length 735;
Best Local Similarity 28.6%; Pred. No. 8.9e-15;
Matches 80; Conservative 12; Mismatches 100; Indels 88; Gaps 16;

QY 248 EHENINGVPHLG----HPFHWGHERSSTTKPPKPHG----SRDHHHPKPHGHPPP 299
||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 408 EHHGTHHSPAHHGEHGTTHHGHGHHHAPAHGHGHSHHSHHSAHHGHGHHHAPAH 467

QY 300 PDERDHSHPPLPQGPPLLPMSCSQCHATFTGNAQASHNNSSDLHPKHSHSEQ 359
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 468 HGHGHHHAP-----AHGHGHHGHTHHGHGSHHSPAH-HGHGHHH 510

QY 360 HPHGHHPH-----AHHPH--EHDTHRQHPHHPH---HGHHPHGHHP 396
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 511 HAPAHGHGHHGHTHHGHGHHHAPAHGHGHGHTHHGHGSHHSPAHHGHGHHH 570

QY 397 ---HGHHPHG-----HHPCHDFQDYGCDPPPHNQHCCHG-HGPPPHLRRRPGKGP 447
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 571 PAHHGHGHHGSHGVHGHG-HESHGHHHAPAH--GH--HGERGVHGH---HGAGYGA 622

QY 448 RPFHCRQIGSVYRLPLRKGEVLPLPEANFFSPFLPHKH 487
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 623 ---HHGHGHHH-----HAPHHEH 639

RESULT 10
Q26056 PRELIMINARY; PRT; 140 AA.
ID Q26056:
AC Q26056:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH PROTEIN (FRAGMENT).
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86174893; PubMed=3007981;
RA Irving D.O., Cross G.A.M., Feder R., Wallach M.;
RT "Structure and organization of the histidine-rich protein gene of
RT Plasmodium lophurae."
RL Mol. Biochem. Parasitol. 18:223-234(1986).
DR EMBL; M15317; AAA29616.1; -;
```

```
DE Y39B6B.GG PROTEIN (Y39B6A.PP PROTEIN).
GN Y39B6B.GG OR Y39B6A.PP.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Sulston J.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132896; CAB60938.1; -;
DR EMBL; AL132948; CAC51077.1; -;
DR HSSP; P13231; IHCE.
DR InterPro; IPR002395; Kininogen.
DR PRINTS; PR00334; KININOGEN.
SQ SEQUENCE 735 AA; 80255 MW; 6EF7B831DBE41159 CRC64;

Query Match 8.9%; Score 269; DB 5; Length 735;
Best Local Similarity 28.6%; Pred. No. 8.9e-15;
Matches 80; Conservative 12; Mismatches 100; Indels 88; Gaps 16;

QY 248 EHENINGVPHLG----HPFHWGHERSSTTKPPKPHG----SRDHHHPKPHGHPPP 299
||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 408 EHHGTHHSPAHHGEHGTTHHGHGHHHAPAHGHGHSHHSHHSAHHGHGHHHAPAH 467

QY 300 PDERDHSHPPLPQGPPLLPMSCSQCHATFTGNAQASHNNSSDLHPKHSHSEQ 359
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 468 HGHGHHHAP-----AHGHGHHGHTHHGHGSHHSPAH-HGHGHHH 510

QY 360 HPHGHHPH-----AHHPH--EHDTHRQHPHHPH---HGHHPHGHHP 396
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 511 HAPAHGHGHHGHTHHGHGHHHAPAHGHGHGHTHHGHGSHHSPAHHGHGHHH 570

QY 397 ---HGHHPHG-----HHPCHDFQDYGCDPPPHNQHCCHG-HGPPPHLRRRPGKGP 447
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 571 PAHHGHGHHGSHGVHGHG-HESHGHHHAPAH--GH--HGERGVHGH---HGAGYGA 622

QY 448 RPFHCRQIGSVYRLPLRKGEVLPLPEANFFSPFLPHKH 487
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 623 ---HHGHGHHH-----HAPHHEH 639

RESULT 10
Q26056 PRELIMINARY; PRT; 140 AA.
ID Q26056:
AC Q26056:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH PROTEIN (FRAGMENT).
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86174893; PubMed=3007981;
RA Irving D.O., Cross G.A.M., Feder R., Wallach M.;
RT "Structure and organization of the histidine-rich protein gene of
RT Plasmodium lophurae."
RL Mol. Biochem. Parasitol. 18:223-234(1986).
DR EMBL; M15317; AAA29616.1; -;
```

RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zavertl J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; The genome sequence of <i>Drosophila melanogaster</i> .;
FT	RL Science 287;2185-2195(2000).
DR	EMBL; AE003716; AAP55377.1; -.
DR	FlyBase; FBgn0038468; CG5225.
DR	InterPro; IPR000087; Collagen.
DR	InterPro; IPR002965; P-rich_extensn.
DR	Pfam; PF01391; Collagen.1.
DR	PRINTS; PR01217; PRIGHXTENS.
DR	SEQUENCE 594 AA; 63898 MW; A747A02C0771C3P CRC64;
Qy	Query Match 8.4%; Score 256; DB 5; Length 594; Best Local Similarity 24.7%; Pred. No. 9.1e-14; Matches 78; Conservative 15; Mismatches 103; Indels 120; Gaps
Db	246 PQEHNTNGVPHLGHFFHMGHERSTTKPFKPGSGSDNNHHNPKHPHEHG-----297     : :     :
Db	132 PGCHKSGHNDHDHNHNHRAPPPPPPPPPPPPPPPPPPPPPPSHSNHRRPPIIVTTPPII 191     : :     :
Qy	298 -PPPDPDERDHS-----GPELPQGPPPLPMSCSQHATF-----332     : :     :
Db	192 PIPLPQKGEGHHNNHKSKGPFGPGPGCTGTCPPGGTTYPQPPPPPPRAPP 251     : :     :
Qy	333 -----GTNGAQRISSHNNSSDLHPKH-----SH 357     : :     :
Db	252 SYPYPPYPPGPGPWIPLVVPWPSPKDGKH-KGDKGDKGHHNYPPGH 310     : :     :
Qy	358 EQ-----HPGHNRHANRP-----HEHD-----THROHPHG- 383     : :     :
Db	311 DKGVYPPRHGGHGNHGRHYPCGYPPPNHPRDGBGKYHNPNNHNGHNHG 370     : :     :
Qy	384 ---HHP---HGNHRPHGNHRPHGNHRCDFQDYCPDCPPRH-----NQHCRCRGHP 433     : :     :
Db	371 HNHHPPHGENHNHPPHNGHGHNHPSHNGGDDPHHTPTGTHCGNGCHTSHQ 430     : :     :
Qy	434 PGGHLRRRGCKGRP 449     : :     :
Db	431 HPHHQ-----PKP 439     : :     :
RESULT	12
Q9K187	ID Q9K187 PRELIMINARY; PRT; 294 AA.
AC	O9K187
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DEF	PUTATIVE VRB (FRAGMENT).
GN	VRB.
OS	Bacillus cereus.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1396;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-AFC 43881;
RX	MEDLINE=20327579; PubMed=10869077;
RA	Schupp J.M., Klevytska A.M., Zniser G., Price L.B., Keim P.;
RT	"vrb", a hypervariable open reading frame in bacillus anthracis.";
J	J. Bacteriol. 182:3989-3997(2000).
DR	EMBL; AF238888..AAF86203.1; -.
DR	InterPro; IPR002395; Kininogen.
DR	PRINTS; PR00334; KININOGEN.
FT	NON_TER 1 SEQUENCE 294 AA; 30753 MW; C05B7B37FE2E395C CRC64:



[illegible]

WJGJ.5  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BRISTOL N2;  
RC MEDLINE=94150718; PubMed=7906398;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Willson R.; Ainscough R.; Anderson K.; Baynes C.; Berks M.;



```

!!AA_SEQUENCE 1.0
ID   HRG_HUMAN          STANDARD;          PRT;          525 AA.
AC   P04196;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histidine-rich glycoprotein precursor (Histidine-proline rich
DE   glycoprotein) (HPRG).
GN   HRG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
[1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86216149; PubMed=3011081;
RA   Koide T., Foster D.C., Yoshitake S., Davie E.W.;
RT   "Amino acid sequence of human histidine-rich glycoprotein derived
RT   from the nucleotide sequence of its cDNA.";
RL   Biochemistry 25:2220-2225(1986).
[2]
RP   SEQUENCE FROM N.A.
RA   Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP   SEQUENCE OF 214-247 FROM N.A.
RX   MEDLINE=94245171; PubMed=8188234;
RA   Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,
RA   Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;
RT   "Evidence for the absence of intron H of the histidine-rich
RT   glycoprotein (HRG) gene: genetic mapping and in situ localization of
RT   HRG to chromosome 3q28-q29.";
RL   Genomics 19:195-197(1994).
[4]
RP   SEQUENCE OF 19-27.
RC   TISSUE=Plasma;
RX   MEDLINE=93092937; PubMed=1459097;
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992)
CC   -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
CC   HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
CC   FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
CC   AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS
CC   HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
CC   MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
CC   COAGULATION CASCADE.
CC   -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
CC   PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATITIONS OF A 5-
CC   RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.
CC   -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M13149; AAA52694.1; -
DR   EMBL; AB005803; BAA21613.1; -
DR   EMBL; Z17218; CAA78925.1; -
DR   PIR; A01287; KGHUGH
DR   SWISS-2DPAGE; P04196; HUMAN.
DR   MW; 142640; -
DR   InterPro; IPR000010; Cystatin.
DR   Pfam; PF00031; cystatin; 1.
DR   SMART; SM00043; CY; 2.
KW   Glycoprotein; Heparin-binding; Repeat; Signal.
FT   CHAIN              19 525
FT   DOMAIN            19 136
FT   DOMAIN            137 254
FT   DOMAIN            276 321
FT   DOMAIN            350 497
FT   DISULFID          24 504
FT   DISULFID          78 89
FT   DISULFID          105 126
FT   DISULFID          203 417
FT   DISULFID          218 241
FT   CARBOHYD           63 63
FT   CARBOHYD           87 87
FT   CARBOHYD          125 125
FT   CARBOHYD          344 344
FT   CARBOHYD          345 345
SQ   SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;
P04196 Length: 525 May 31, 2002 14:35 Type: P Check: 9896 ..
1  MKALIAALL ITLOYSCAVS PTDSCAVEPE AEKALDLINK RRDGYLFQL
51  LRIADAHLDL RENTTVYVLV LDVQSDCSV LSRKYWDCE PPSRRPSEI
101 VIGQCKVIAT RHSHESQDLR VIDFNCTTSS VSSALANTKD SPVLIDFFED
151 TERYRKQANK ALEKYKEEND DFASFVRVDRI ERVARVRGGE GTGYFVDFSV
201 RNCPRHFFPR HPNVFGFCRA DLFYDVEALD LESPKNLVIN CEVFDPOEHE
251 NINGVPPHLG HPFWGGHER SSTTKPPFKP HGSRDHHHPH KPHEHGPPPPP
301 PDERDHSRGP PLPGQPPPLL PMSCSSCQHA TFGTNGAQRH SHNNSSDLH
351 PKHHHSHEQH PGHHHPHAHH PHEHDTHQH PGHHHPHGHH PHGHHHPGHH
401 PGHHHPCHD FQYGGCDPP PHNQGHCHG HGPPPGHLR RGPGKGRPF
451 HCRQTGSYVR LPPLRKGEVL PLPEANFPSP PLPHHKKHLK PDNQFPQSV
501 SESCPGKFKS GFPQVSMFFT HTPPK
HISTIDINE-RICH GLYCOPROTEIN.
CYSTATIN-LIKE 1.
CYSTATIN-LIKE 2.
PRO-RICH.
PRO/HIS-RICH.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
A2B124D6CE93114F CRC64;

```

**THIS PAGE BLANK (USPTO)**